DOCUMENT RESUME

ED 044 611 AC 008 771

AUTHOR Cain, Glen G.; Hollister, Robinson G.
TITLE The Methodology of Evaluating Social Action Programs.
INSTITUTION Wisconsin Univ., Madison. Center for Studies in Vocational and Technical Education.
PUB DATE Nov 69
NOTE 33p.; Reprinted from Public-Private Manpower Policies
AVAILABLE FROM Industrial Relations Research Institute, 4315 Social Science Building, 1180 Observatory Drive, Madison, Wisconsin 53706

EDRS PRICE EDRS Price MF-$0.25 HC-$1.75
DESCRIPTORS *Action Research, Bibliographies, Cost Effectiveness, *Evaluation Criteria, Manpower Development, *Program Evaluation, *Research Methodology, *Social Action, Statistical Analysis

ABSTRACT

Focusing on the evaluation of outcomes ("cost benefit analysis") in large scale social action programs, this paper examines issues relating to the adequacy of theory and methodology as well as the impact of different types of persons (academics, politicians, program administrators) involved in the evaluation process. Problems of evaluation design--control groups, the criterion of replicability (applicability on a wider scale), statistical analysis, and implications of differing socioeconomic criteria--are considered in detail. The authors then propose a deliberately experimental approach which would permit program planners and administrators to learn faster by testing alternative concepts simultaneously. Using an analogy with the court system, they also discuss the potential value of a "rules in evidence" approach to setting standards for acceptance of evaluation results. The document includes 19 footnotes, followed by reprints and other publications by the Industrial Relations Research Institute and the Center for Studies in Vocational and Technical Education. (LY)

THE UNIVERSITY OF WISCONSIN

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

CENTER FOR STUDIES IN VOCATIONAL AND TECHNICAL EDUCATION

The Methodology of Evaluating Social Action Programs

ED 044611

by Glen G. Cain and Robinson G. Hollister

Reprinted from
Public-Private Manpower Policies
Industrial Relations Research Association
November 1969

AC008771

INDUSTRIAL RELATIONS RESEARCH INSTITUTE

INDUSTRIAL RELATIONS RESEARCH INSTITUTE
THE UNIVERSITY OF WISCONSIN

The Industrial Relations Research Institute fosters and coordinates study and research in the field of industrial relations throughout the various schools and departments of The University of Wisconsin. In carrying out these activities, the University continues a tradition of more than half a century of leading work in this field.

The Institute's programs of graduate study, leading to Master's and Ph.D. degrees in Industrial Relations, are closely associated with course and seminar offerings in the departments of economics, history, political science, psychology, and sociology in the College of Letters and Science, and with those in the College of Engineering and the Schools of Business, Education, and Law.

Faculty members participating in the research, teaching, and coordinating functions of the Institute are also drawn from these schools and departments. Current committee members include:

Executive Committee

Chairman, James Stern, Economics; Director I.R.R.I.
Jack Barbash, Economics
Leonard Berkowitz, Psychology
Abner Brodie, Law
Murray Edelman, Political Science
Stephen C. Kleene (ex officio), Dean, L.&S.
Nathan P. Feinsinger, Law
Alton Johnson, Business
Richard U. Miller, Business;
Associate Director, I.R.R.I.
Raymond Munts, Social Work
Gerald Nadler, Industrial Engineering
Robert Ozanne, School for Workers
Charles Perrow, Sociology
Morton Rothstein, History
Gerald G. Somers, Economics
George Strothar, Vice Chancellor, University Extension
Carl Schramm, Student

Admissions Committee

Chairman, Richard U. Miller, Business
Arlen Christenson, Law
David Gustafson, Engineering
Solomon B. Levine, Economics and Business
Eleanore Roe, Law
Donald Schwab, Business
Lee Dyer, Student

Curriculum Committee

Chairman, David B. Johnson, Economics
Charles Bridgman, Psychology
Alan Filley, Business
Donald Treiman, Sociology
Roy Adams, Student
Herbert G. Heneman, III, Student
Sanford Sherman, Student
John Baum, Student

Publications Committee

Chairman, Gerald G. Somers, Economics
L. L. Cummings, Business
Scott Cutlip, Journalism
Barbara Dennis, *Journal of Human Resources*
William Glade, Business and Economics
George Hagglund, School for Workers
Everett Kassalow, Economics
Jack Ladinsky, Sociology
Thomas Decotiis, Student

Research Committee

Chairman, Solomon B. Levine, Economics and Business
Andre Delbecq, Business
Alan Gross, Psychology
Richard Hamilton, Sociology
James E. Jones, Jr., Law
Hervey Juris, School for Workers
Robert J. Lampman, Economics
David Zimmerman, Student

ED044611

Reprinted from PUBLIC-PRIVATE MANPOWER POLICIES
Industrial Relations Research Association

The Methodology of Evaluating Social Action Programs

By Glen G. Cain * and Robinson G. Hollister *

Apologia

This paper is largely motivated by our experiences as academics who became directly enmeshed in the problems of a public agency which was under considerable pressure—generated by both the agency staff itself and external factors—to "evaluate" manpower, and other social action, programs.

It became evident that there were several major obstacles to effective evaluation in this context. These obstacles were created both by the several types of "actors" necessarily involved in such evaluation efforts and by complications and weaknesses in the theory and methodology to be applied. Difficulties of communication among the "actors," due both to differences in training and to suspicions about motives, often made it hard to distinguish between difficulties arising because the theory was weak and those arising because adequate theory was poorly understood.

In this paper we try to separate out some of these issues, both those concerning the adequacy of theory and methodology and

---

* This research was supported by funds granted to the Institute for Research on Poverty, University of Wisconsin, pursuant to the provisions of the Economic Opportunity Act of 1964. Professor Cain and Professor Hollister are associated with the University of Wisconsin Department of Economics and are members of the Institute staff. The authors are grateful to the following persons, who have increased their understanding of the ideas in this paper or have commented directly on an earlier draft (or have done both): David Bradford, Frank Cassell, John Evans, Woodrow Ginsburg, Thomas Glennan, Robert Levine, Guy Orcutt, Gerald Somers, Ernst Stromsdorfer, Harold Watts, Arnold Weber, Burton Weisbrod, and Walter Williams. A longer version of this paper is available as Discussion Paper 42-69 from the Institute for Research on Poverty, University of Wisconsin, Madison. An intermediate length version will appear in the volume consisting of the Proceedings of the North American Conference on Cost-Benefit Analyses, held in Madison, Wisconsin, May 14–15, 1969.

those relating to the various sorts of actors. We have sought to couch the discussion in language that will make it available to academics, who we feel need a heightened awareness of the more practical difficulties of execution of evaluations in the social action context—and to public agency and political personnel, who we believe would benefit from increased sensitivity to the ways in which careful consideration of the design and careful control of evaluations can increase the power of the information derived from sueh efforts. The attempt to reach both audiences in one paper produces a mixture of elements bound to strike members of either audience as, at some points, extremely naive and, at others, disturbingly recondite. We can only hope that such reactions will be transformed into a resolve to initiate a more meaningful dialogue on these issues, a dialogue we feel is crucial to the development of an effective approach to evaluations of social action programs.

Introduction

This paper began as a discussion of methods of evaluating manpower programs—programs which used to consist almost entirely of vocational training and various but limited types of assistance for the worker searching for jobs within local labor markets. But with the recent emphasis on problems of poverty and the disadvantaged worker, manpower programs have come to involve remedial and general education, to intermesh with community action programs providing a variety of welfare services, and, on a trial basis, to assist in migration between labor markets. They are part of a broader class of programs which, for lack of a better term, we might call social action programs. Our paper will include many references to this broader class, and in particular to anti-poverty programs. In so doing, we hope to provide a more general and more relevant perspective on the topic of evaluation methodology.

We hold the opinion, apparently widely shared, that existing evaluations of social action programs, (and we are including our own), have fallen short of meeting the standards possible within

the disciplines of the social sciences. The reasons for these shortcomings are easy to identify. The programs typically involve investments in human beings, a relatively new area of empirical research in economics. They are aimed at such social and political goals as equality and election victories, as well as economic objectives concerning, say, income and employment. They often attempt to deliver services on a large enough scale to make a noticeable impact upon the community. And at the same time, they are expected to provide a quasi-experimental basis for determining what programs ought to be implemented and how they ought to be run.

It is not surprising, then, that evaluations of social action programs have often not been attempted and when attempted, have not been successful. Despite this background, we believe that existing data and methods permit evaluations which, while not satisfying the methodological purists, can at least provide the rules of evidence for judging the degree to which programs have succeeded or failed. Specifically, the theme we will develop is that evaluations should be set up to provide the ingredients of an experimental situation: a model suitable for statistical testing, a wide range in the values of the variables representing the program inputs, and the judicious use of control groups.

The paper reflects several backgrounds in which we have had some experience—from economics, the tradition of benefit-cost analyses; from the other social sciences, the approach of quasi-experimental research; and from a governmental agency, the perspective of one initiating and using evaluation studies. Each of these points of view has its own literature which we have by no means covered, but to which we are indebted.[1]

Types of Evaluation

There are two broad types of evaluation. The first, which we call "process evaluation," is mainly administrative monitoring. Any program must be monitored (or evaluated) regarding the integ-

rity of its financial transactions and accounting system. There is also an obvious need to check on other managerial functions, including whether or not accurate records are being kept. In sum, "process evaluation" addresses the question: Given the existence of the program, is it being run honestly and administered efficiently?

A second type of evaluation, and the one with which we are concerned, may be called "outcome evaluation," more familiarly known as "cost-benefit analysis." Although both the inputs and outcomes of the program require measurements, the toughest problem is deciding on and measuring the outcomes. With this type of evaluation the whole concept of the program is brought into question, and it is certainly possible that a project might be judged to be a success or a failure irrespective of how well it was being administered.

A useful categorization of cost-benefit evaluations draws a distinction between a priori analyses and ex post analyses. An example of a priori analysis is the cost-effectiveness studies of weapons systems conducted by the Defense Department, which have analyzed war situations where there were no "real outcomes" and, thus, no ex post results with which to test the evaluation models. Similarly, most evaluations of water resource projects are confined to alternative proposals where the benefits and costs are estimated prior to the actual undertaking of the projects.[2] Only in the area of social action programs such as poverty, labor training, and to some extent housing, have substantial attempts been made to evaluate programs, not just in terms of before-the-fact estimates of probable outcomes or in terms of simulated hypothetical outcomes, but also on the basis of data actually gathered during or after the operation of the program.

A priori cost-benefit analyses of social action programs can, of course, be useful in program planning and feasibility studies, but the real demand and challenge lies in ex post evaluations. This more stringent demand made of social action programs may say

something about the degree of skepticism and lack of sympathy Congress (or "society") has concerning these programs, but this posture appears to be one of the facts of political life.

Problems of the Design of the Evaluation [2A]

A. The Use of Control Groups

Given the objective of a social action program, the evaluative question is: "What difference did the program make?", and this question should be taken literally. We want to know the difference between the behavior with the program and the behavior if there had been no program. To answer the question, some form of control group is essential. We need a basis for comparison—some base group that performs the methodological function of a control group. Let us consider some alternatives.

The Before-and-After Study. In the before and after study, the assumption is that each subject is his own control (or the aggregate is its own control) and that the behavior of the group before the program is a measure of performance that would have occurred if there had been no program. However, it is well known that there are many situations in which this assumption is not tenable. We might briefly cite some examples found in manpower programs.

Sometimes the "before situation" is a point in time when the participants are at a particularly low state—lower, that is, than is normal for the group. The very fact of being eligible for participation in a poverty program may reflect transitory conditions. Under such conditions we should expect a "natural" regression toward their mean level of performance if we measure their status in an "after situation," even if there were no program in the intervening period. Using zero earnings as the permanent measure of earnings of an unemployed person is an example of attributing normality to a transitory status.

Another similar situation arises when young people are involved in the program. Ordinary maturation and the acquisition

of experience over the passage of time would be expected to improve their wages and employment situation.

There may be some structural change in the personal situations of the participants before and after the program, which has nothing to do with the program but would vitiate any simple before-or-after comparison. We should not, for example, look upon the relatively high earnings record of coal miners or packinghouse workers as characteristic of their "before situation" if, in fact, they have been permanently displaced from their jobs.

As a final example of a situation in which the before-and-after comparison is invalid, there is the frequent occurrence of significant environmental changes—particularly in labor market environments—which are characterized by seasonal and cyclical fluctuations. Is it the program or the changed environment which has brought about the change in behavior? All of the above examples of invalidated evaluations could have been at least partially corrected if the control groups had been other similar persons who were in similar situations in the pre-training period.

Control Groups and Small Group Studies. The particular strength of the small scale study is that it greatly facilitates the desideratum of random assignments to "treatment groups" and "control groups" or, at least, a closely supervised matching of treatment and control groups. Its particular shortcoming is that it is likely to lack representativeness—both in terms of the characteristics of the program participants and in terms of the character of the program. There is first the problem of a "hot house environment" of the small group study. (See discussion of "replicability" below.) Second, a wide range of values of the program inputs (i.e., in terms of levels of a given treatment or in terms of qualitatively different types of treatments) is less likely to be available in a small group study. Third, the small group study may not be able to detect the program's differential effects on different types of participants (e.g., by age, sex, color, residence, etc.,) either because the wide variety of participant types are not available or

because their numbers are too small. Finally, it is both a strength and a weakness of the small scale study that it is usually confined to a single geographic location. Thus, although "extraneous" noise from different environments are eliminated, we may learn little or nothing about how the program would operate in different environments.

Control Groups and Large Group Studies. The large scale study, which involves gathering data over a wide range of environments, customarily achieves "control" over the characteristics of participants and nonparticipants and over programs and environmental characteristics by statistical methods, rather than by randomization or careful matching, individual by individual. These studies have the capability of correcting each of the shortcomings attributed to the small scale studies in the preceding paragraph. But because they are almost impossible to operate with randomization, the large scale studies run afoul of the familiar problem in which the selectivity of the participants may be associated with some unmeasured variable(s) which makes it impossible to determine what the net effect of the treatment is. Since this shortcoming is so serious in the minds of many analysts, particularly statisticians, and because the small scale studies have a longer history of usage and acceptability in sociology and psychology, it may be worthwhile to defend at greater length the large scale studies, which are more common to economists.

Randomization is seldom attempted for reasons having to do with the attitudes of the administrators of a program, local pressures from the client population, or various logistic problems. Indeed, all these reasons may serve to botch an *attempted* randomization procedure. Furthermore, we can say with greater certitude that the ideal "double-blind experiment with placebos" is almost impossible to achieve. If we are to do something other than abandon evaluation efforts in the face of these obstacles to randomization, we will have to turn to the large scale study and the statistical design issues that go along with it.

The fact that the programs vary across cities or among administrators may be turned to our advantage by viewing these as "natural experiments" [3] which may permit an extrapolation of the results of the treatment to the "zero" or "no-treatment" level. The analyst should work with the administrator in advance to design the program variability in ways which minimize the confounding of results with environmental influences. Furthermore, ethical problems raised by deliberately excluding some persons from the presumed beneficial treatments are to some extend avoided by assignments to differing treatments (although, here again, randomization is the ideal way to make these assignments).

It is difficult at this stage, to provide more than superficial observations regarding the choice between small and large-scale studies. It would seem that for those evaluations that have a design concept which is radically different from existing designs or where there is a quite narrow hypothesis which requires detailed examination, a small group study would be preferable. Conversely, when the concept underlying a program is quite broad and where large amounts of resources are to be allocated, the large group approach is probably more relevant—a point argued in greater detail in our discussion of the "replicability criterion."

B. The Replicability Criterion

A source of friction between administrators of programs and those doing evaluation research, usually academicians, is the failure to agree upon the level of decision-making for which the results of the evaluation are to be used. This failure, which is all the more serious because the issue is often not explicitly addressed, leads to disputes regarding two related issues—the scope of the evaluation study and the selection of variables to be studied. To deal with these disputes, we suggest applying the "replicability criterion." We apply this name to the criterion because of the large number of cases in which evaluations of concepts have been made on the basis of projects which are not likely to be replicable on a large

scale or which focus on characteristics of the project which are not within the ability of the decision-makers to control. To take an extreme example, it has sometimes been stated that the success of a compensatory education program depended upon the "warmth and enthusiasm" of the teachers. In the context of a nationwide program, no administrator has control over the level of "warmth and enthusiasm" of teachers.

It is sometimes argued by administrators that evaluations which are based upon samples drawn from many centers of a program are not legitimate tests of the program concept since they do not adequately take into account the differences in the details of individual projects or of differentiated populations. These attitudes frequently lead the administrators or other champions of the program to select, either ex ante or ex post, particular "pet" projects for evaluations that "really count." In the extreme, this approach consists of looking at the successful programs (based on observations of ongoing or even completed programs) and then claiming that these are really the ones that should be the basis for the evaluation of the program as a whole. *If* these successful programs have worked with representative participants in representative surroundings and *if* the techniques used—including the quality of the administrative and operational personnel—can be replicated on a nationwide basis, *then* it makes sense to say that the evaluation of the particular program can stand for an evaluation of the overall program. But we can seldom assume these conditional statements. After all, each of the individual programs, a few political plums notwithstanding, was set up because someone thought it was worthwhile. Of course, some will flop because of poor teachers or because one or more operations were fouled up—but it is in the nature of the beast that some incompetent administrative and operational foul-ups will occur. A strength of summary, over-all measures of performance is that they will include "accidental" foul-ups with the "accidental" successes, the few bad administrators and teachers as well as the few charismatic leaders.

As a case in point, consider the success (according to prevailing opinion) of Reverend Sullivan's Operation Industrial Council in Philadelphia with the (as yet) absence of any evidence that the OIC idea has been successfully transferred elsewhere.[4]

Small scale studies of pre-selected particular programs are most useful either for assessing radically different program ideas or for providing the administrator with information relevant to decisions of program content *within* the confines of his overall program. These are important uses, but the decisions at a broader level which concern the allocation of resources among programs of widely differing concepts call for a different type of evaluation with a focus on different variables.

It may be helpful to cite an example of the way in which the replicability criterion should have been applied. A few years ago, a broad scale evaluation of the Work Experience Program[5] was carried out. (The evaluation was of necessity based upon very fragmentary data, but we are here concerned with the issues it raised rather than with its own merits.) The evaluation indicated that on the average the unemployment rates among the completers of the program were just as high as those with similar characteristics who had not been in the program. On the basis of this evaluation, it was argued that the concept of the program was faulty, and some rather major shifts in the design and in the allocation of resources to the program were advocated.[6] Other analysts objected to this rather drastic conclusion and argued that the "proper" evaluative procedure was to examine individual projects within the program, pick out those projects which had higher "success rates," and then attempt to determine which characteristics of these projects were related to those "success rates."[7]

The argument as to which approach is proper depends on the particular decision framework to which the evaluation results were to be applied. To the administrators of the program, it is really the project by project type of analysis which is relevant

to the decision variables which they control. The broader type of evaluation would be of interest, but their primary concern is to adjust the mix of program elements to obtain the best results within the given broad concept of the program. Even for program administrators, however, there will be elements and personnel peculiar to a given area or project that will not be replicable in other areas and other projects.

For decision-makers at levels higher than the program administrator the broader type of evaluation will provide the sort of information relevant to their decision frame. Their task is to allocate resources among programs based upon different broad concepts. Negative findings from the broader evaluation argue against increasing the allocation to the program, although a conservative response might be to hold the line on the program while awaiting the more detailed project-by-project evaluation to determine whether there is something salvageable in the concept embodied in the program. There will always be alternative programs serving the same population however, and the decision-maker is justified in shifting resources toward those programs which hold out the promise of better results.

The basic point is that project-by-project evaluations are bound to turn up some "successful" project somewhere, but unless there is good evidence that that "success" can be broadly replicated and that the administrative controls are adequate to insure such replication, then the individual project success is irrelevant. Resources must be allocated in light of evidence that concepts are not only "successful" on a priori grounds or in particular small-scale contexts but that they are in fact "successful" in large-scale implementation.

C. The Theoretical Framework—Some Statistical Considerations.

The main function of a theoretical framework in cost-benefit evaluations is to provide a statistical model suitable for testing.

In this section a few brief remarks will be made about the statistical design of the evaluation—a lengthier discussion of these matters is taken up in another paper.[7A] In these remarks we will adopt the terminology of regression analysis, which is a statistical method flexible enough to handle an analysis of variance approach or that involved in simply working with cell values in tables. In the regression model, the dependent variable is the objective of the social action program and the particular set of independent variables of most interest to us are those that describe or represent the program, or program inputs. In this discussion the independent variables will sometimes be referred to as "treatment variables."

It may be useful to divide the problems of statistical design into two categories: First, attaining acceptable levels of statistical significance on the measured effects of the treatment variables; second, measuring those effects without bias. We will not discuss the first problems here except to note that the failure to attain statistical significance of the effect of the treatment variable occurs either because of large unexplained variation in the dependent variable or small effects of treatment variables and these can be overcome with sufficiently large sample sizes. In our opinion, the most serious defect in evaluation studies is biases in the measures of effects of the treatment variables, and this error is unlikely to be removed by enlarging the sample size.

One source of bias is inaccurate measures of the treatment variable, but a more pervasive and more serious problem is the presence of variables, not included in the statistical model, which are correlated with both the dependent variable and the treatment variable. Had the assignment to a program been made on a random basis, the laws of probability would have assured a low correlation (zero in the limit of a large enough sample size) between participation in the program and these omitted variables. In the absence of randomization, we must fall back on statistical controls. At this point our theory and a priori information are crucially im-

portant. The requirements are obvious: to identify the variables whose omission leads to biases in the measured effects of the treatment variables and to include them in the model. These variables may be objectively measurable, such as age or education or previous work experience. Or they may be such difficult-to-measure characteristics as ambition, motivation, or an "appealing personality." [8]

As we know too well, however, our theories are woefully weak in providing us with the correct list of variables for explaining such dependent variables as income change, employment experience, health status, or educational attainment, and we often do not have measures of those we do know about. The latter problem frequently arises because of the unfortunate practice of inviting the evaluator in *after* the program has been run and the data have been collected. Even in the best of situations regarding the availability of objective measures of important variables, if we do not have random assignments we must still admit the possibility that *self-selectivity* or the *selectivity procedures* of the program administrators has introduced a systematic difference between the participants and the nonparticipants. We do not claim, as the purists would, that nonrandom procedures invalidate all evaluations, although there are cases when they undoubtedly have, but the advantages of randomization are immense and we can do a great deal more to achieve this procedure if we can only convince each other of its importance.

Another important advantage of randomization should be mentioned. We have noted that variables which are correlated with both the treatment variable and the dependent variable must be included in the model to measure treatment effects without bias. However, since our information about the effect of the treatment variable necessarily depends on variability in treatments, and since the only variation we can observe within the framework of the statistical model is the residual variation in treatments—that is, variation which remains after the entire set

of independent variables is included, greater efficiency is obtained when the treatment variable is uncorrelated with the other independent variables. In the opposite extreme, if the treatment variables were perfectly correlated with some other variable or combination of variables, we would be unable to distinguish between which of the two sets of factors caused a change. It follows that even in the absence of randomization, designing the programs to be studied with as wide a range in levels and types of "treatments" as possible will serve to maximize the information we can extract from an ex post analysis.

There are reasons in addition to those of statistical efficiency for planning for a wide range of values in the treatment of programmatic variables. One is that social action programs have a tendency to change, rather frequently and radically, during the course of their operation. Evaluations designed to test a single type of program are rendered meaningless because the program-type perishes. But if the design covers a wider variety of programs, then a built-in hedge against the effects of change is attained. Indeed, there is an even more fundamental reason why a wide range of inputs and program types should be planned for, and it is simply this: we seldom know enough about what will work in a social action program to justify putting our eggs in the single basket of one type of program. This evaluation model for a single type of project, sometimes described as the analogue of the "pilot plant," is not the appropriate model for social action programs given our current state of knowledge.[9]

D. The Theoretical Framework—Some Economic Considerations.

For operational purposes we will assume that the evaluation of each social action program can, at least in principle, be cast in the statistical model discussed in the previous section, complete with variables representing an objective of the program, treatment variables representing the program inputs, control variables, and con-

trol groups.[10] However, the substantive theoretical content of these models—the particular selection of variables and their functional form—must come from one or more of the traditional disciplines such as educational psychology (e.g., for Head Start), demography (e.g., for a family planning program), medical science (e.g., for a neighborhood health center), economics (e.g., for a manpower training program), and so on.

Sooner or later economics must enter all evaluations, since "costing out" the programs and the setting of implicit or explicit dollar measures of the worth of a program are essential steps in a complete evaluation. In making the required cost-benefit analysis, the part of economic theory that applies is the investment theory of public finance economics, with its infusion of welfare economics. The function of investment theory is to make commensurable inputs and outcomes of a social action program which are spaced over time.[10a] Welfare economics analyzes the distinctions between financial costs and real resource costs, between direct effects of a program and externalities, and between efficiency criteria and equity (or distributional) criteria.

We will say very little on the last mentioned distributional or equity question of *who pays* and *who recieves*, even though we strongly feel that accurate data on the distribution of benefits and costs is essential to an evaluation of social action programs. However, the task of conducting a "conventional" benefit-cost analysis (where the criterion is allocative efficiency) is sufficiently complex that we believe it preferable to separate the distributional questions.

Program Inputs. In the investment theory model costs are attached to all inputs of a program and a single number emerges which measures the present value of the resources used. Most of the technical problems faced by the analysts on the input side are those of traditional cost accounting. We will confine our remarks to the two familiar and somewhat controversial problems of opportunity costs and transfer payments, which arise in nearly every manpower program. Both of these problems are most effectively

dealt with if one starts by asking: What is the decision context for which these input measures are defined?

The most general decision context—and the one to which economists most naturally refer—is that of the productivity of alternative resources utilizations in society or the nation *as a whole.* In this case, one wishes to measure the cost of inputs in terms of the net reduction in value of alternative socially productive activities caused by the use of the inputs in this particular activity. Now, the value of most inputs in terms of their alternative use will be more or less clearly indicated by their market price, but there are some inputs for which this will not be true. The most troublesome cases often concern the time of people. A well known example is the value of the time spent by students in school: since those over 14 or so could be in the job market, the social product (or national income) is less; therefore, an estimate is needed of what their earnings would be had they not been in school. (Such an estimate should reflect whatever amount of unemployment would be considered "normal.") For manpower programs the best evaluation design would provide a control group to measure the opportunity costs of the time spent by the trainees in the program.

Sometimes the prices of inputs (market prices or prices fixed by the government) do not adequately reflect their marginal social productivity, and "corrected" or "shadow prices" are necessary. For example, the ostensible prices of leisure or of the housework of a wife are zero and obviously below their real price. By contrast a governmental fixed price of some surplus commodity is too high.

The definition and treatment of transfer payments also depend on the decision context of the analysis. From the national perspective money outlays from the budget of one program that are offset by reduced outlays elsewhere in society do not decrease the value of the social product. When these outlays are in the form of cash payments or consumption goods, they are called transfer payments. An example is the provision of room and board for Job Corps

trainees. Since it must be assumed that someone (their parents, themselves, or some welfare agency) would be meeting the costs of their room and board if they were not in the program, the provision of these services by the program reflects no *net* reduction in the value of alternative socially productive activities. Whoever was paying these costs before will be relieved of that burden and will spend the money thus saved on other goods and services. If there has been an actual *increase* in the value of food consumed by the trainee or in the quality of his housing, the net increase can be counted as a program input—a cost. But in general, it would be equal to the net increase in the value of food and housing consumed—a benefit.[11] To summarize, if these input costs are simply being *transferred* from one individual or agency to another individual or agency they either represent no real cost of resources of this program or they are a cost which is immediately offset by the benefit it yields to the recipient—remembering that the decision context is the general one which includes all members of society, with no one member receiving any different weight in the calculation of benefits.

In a narrower decision context, the accounting basis may shift; some input costs counted in the broader context are not counted in the narrower one and vice versa. One example of a narrow decision context—a favorite of people in government, but repugnant to most economists—is the vaguely defined "public budget." Alternatively the decision context might be considered that of the "taxpayers' viewpoint" if the program participants and their families are excluded from the group considered as taxpayers. In this context the only costs that are to be counted are those that come from the public budget. Some of the examples we discussed above are now reversed. Presumably, most of the opportunity costs of a student's time spent in school is of no interest to the taxpayer since it is a "cost" which is not directly imposed upon the public budget. (A qualification is that the taxpayer should be interested in the taxes the student would pay if he were working). By contrast the pay-

ments for the cost of room and board to a Job Corpsman, which was considered a transfer payment above, would now be considered an input cost from the "taxpayer's viewpoint." The fact that the trainee or his family is relieved of this burden would be of no interest since it would not be reflected in the public budget. However, if the costs of room and board had been met previously by a public welfare agency, then from the "taxpayer's viewpoint," the costs would not be charged to the Job Corps program.

It is not uncommon to see several decision contexts used in one analysis, and used inconsistently. For example, the post-training earnings improvement from participation in a Job Corps program are considered benefits. We all recognize, of course, that the earnings will be used mostly for consumption by the Job Corps graduate. But in the same study, his consumption during training (room, meals, and spending allowance), is not viewed as conferring benefits to the corpsman.[12] Or is it that the benefits should not count because while in training, he is not considered a member of "our society?" We leave this puzzle to those who prefer these restricted decision contexts. There are other such examples and still other and more narrow decision contexts, such as that of a local government or of the project by itself. But it is probably clear that our preference is for the national or total societal perspective.

Program Outcomes. The problems of measurement on the outcome side of the evaluation problem are tougher to handle, and ex post evaluations of social action programs face particular problems because these outcomes are likely to involve behavioral relationships which are not well understood. It is particularly difficult to predict long run or permanent behavioral changes from the short run indicators revealed by the on-going or just completed program.

The outcomes we wish to measure from many social action programs occur months or years after the participants have completed the program. We can use proxy measures, which can themselves be measured during and soon after the program, but follow-up studies are clearly preferred and may in many cases be essential. A good

deal depends on the confidence we have in the power of our theories to link the proxies or short-run effects (e.g., test scores, health treatments, employment experience in the short-run, etc.) with the longer run goals (longer run educational attainment, longevity, incomes, or all of these and perhaps other "softer" measures of "well-being"). It is a role for "basic research" in the social sciences to provide this type of theoretical-empirical information to evaluations, but we can also hope that the more thorough evaluation studies will contribute to our stock of "basic research" findings.

The major obstacle to follow-up measures is the difficulty in locating people, particularly those from disadvantaged populations who may be less responsive and who have irregular living patterns. The biases due to nonresponse may be severe, since those participants who are easiest to locate are likely to be the most "successful," both because of their apparent stability and because those who have "failed" may well be less responsive to requests to reveal their current status. One way around the costly problem of tracking down respondents for earnings data is to use Social Security records for participant and control groups. The rights of confidentiality may be preserved by aggregating the data.

Another problem in measuring outcomes, which also tends to be more talked about despairingly than coped with positively, is the category of external or third-party effects of the program. As a typical illustration consider a youth training program, which not only increases the earnings of the youths, but also reduces the incidence of crime among these groups, which generally benefits the community—e.g. less damage and lower costs of prevention and rehabilitation programs. Another source of third-party effects are those accruing to the participant's family members, including those yet to be born. It is an open question, however, whether the problem for concern is the lack of measurement of these external effects, or the tendency by administrators and others (particularly friends of the programs) to exaggerate their likely importance and to

count as external or secondary benefits those effects which, while benefiting some people do so at the expense of others.[13]

Concerning training and education programs, in particular, two types of effects that have received scant investigation are "negative effects" and those which affect the structure of communities. A discussion, though little measurement, of such effects has appeared in studies and accounts of public housing, urban renewal, and road building programs.[14] The following list of three potential negative effects of manpower programs can serve as examples.

(a) Programs placing the hard-core poor into jobs have had, according to some reports, disruptive effects in the plant—both because of the behavior of the trainee-participants (e.g., disciplinary problems and high rates of absenteeism) and because of the special treatment which the participants received.

(b) Programs which augment the supply of workers in a particular occupation will have the effect of exerting downward pressure on the wages of existing workers in that occupation. It is worth noting that the workers earning high wages are likely to belong to unions which will block these programs in their field (e.g., the building trades), but that low wage workers (like hospital workers) have little or no power to protect their economic interests.

(c) Programs which engender high hopes among some applicants or entrants may lead to a further alienation and hostility for some of those who are rejected or otherwise refused admission or for those who enter and fail. Admission policies are, in fact, just one example of administrative discretionary behavior that can have considerable separate influence on the positive and negative effects of programs—a point brought out in debates about the relative merits of self-help programs, transfer payment programs, and welfare and relief programs.[15]

Community effects of a social action program can be viewed as a special type of external effect, since the changes in the community structure or in various community institutions are assumed to be important because of the benefits or costs they ultimately pro-

vide for third-party individuals in the community. Thus, we are not proposing that the "community" be viewed as an "entity" separate from the individuals who comprise it. However, a separate focus on measures of community institutional changes appears necessary since the present state of our theories of community organization permit us little scope for anything except qualitative linkages between institutional changes and their effects on individuals in the community. We can, for example, consider better communication between the neighborhood populace and the police, school officials, or the employment service as "good things," either in their own right, as expressions of the democratic ethic, or because we believe that such changes will have tangible effects in safety, school achievement or better jobs.

Intentional Experiments: A Suggested Strategy

Underlying the growing interest in evaluations of social action programs is the enlightened idea that the scientific method can be applied to program experience to establish and measure particular cause and effect relationships which are amenable to change through the agents of public policy. However, traditional methods in science, whether the laboratory experimentation of the physical scientists, the testing of pilot models by engineers, or field testing of drugs by medical scientists, are seldom models that can be directly copied, helpful though they are as standards of rigor.

In particular, evaluation designs patterned after the testing of pilot models, which correspond to "demonstration projects" in the field of social action programs, have been inadequate for both theoretical and operational reasons. The present state of our theories of social behavior does not justify settling on a unique plan of action, and we cannot, almost by definition, learn much about alternative courses of action from a single pilot project. It is somewhat paradoxical that on the operational level the pilot model has failed to give us much information because the design has frequently been impossible to control and has spun off in different directions.

The combination of, first, loose administration of and rapid changes in the operation of individual projects and second, a large scale program with many heterogeneous projects (different administrations, different environments, different clientele, etc.), has led to the interesting view that this heterogeneity creates what are, in effect, "natural experiments" for an evaluation design. For economists, who are used to thinking of the measurement of consumers' responses to changes in the price of wheat or investors' responses to changes in the interest rate, the idea of "natural experiments" has a certain appeal. But what should be clear from this discussion—and others before us have reached the same conclusion—is that a greatly improved evaluation could be obtained if social action programs were initiated in *intentional* experiments.

When one talks of "experiments" in the social sciences what inevitably comes to mind is a small scale, carefully controlled study, such as those traditionally employed in psychology. Thus, when one suggests that social action programs be initiated as intentional experiments, people imagine a process which would involve a series of small test projects, a period of delay while those projects are completed and evaluated, and perhaps more retesting before any major program is mounted. This is very definitely *not* what we mean when we suggest social action programs as intentional experimentation. We would stress the word *action* to highlight the difference between what we suggest versus the traditional small scale experimentation.

Social action programs are undertaken because there is a clearly perceived social problem that requires some form of amelioration. In general, (with the exception perhaps of the area of medicinal drugs were a counter tradition has been carefully or painfully built up), we are not willing to postpone large scale attempts at amelioration of such problems until all the steps of a careful testing of hypotheses, development of pilot projects, etc. have been carried out. We would suggest that large scale ameliorative social action and intentional experimentation are not incompatible; experi-

mental designs can be built into a large scale social action program.

If a commitment is made to a more frankly experimental social action program by decision-makers and administrators, then many of the objectives we have advocated can be addressed directly at the planning stage. If we begin a large national program with a frank awareness that we do not know which program concept is more likely to be most efficacious, then several program models could be selected for implementation in several areas, with enough variability in the key elements which make up the concepts to allow good measures of the differential responses to those elements. If social action programs are approached with an "intentionally experimental" point of view, then the analytical powers of our statistical models of evaluation can be greatly enhanced by attempts to insure that "confounding" effects are minimized—i.e., that program treatment variables are uncorrelated with participant characteristics and particular types of environments.

A less technical but equally important gain from this approach to social action programs is the understanding on the part of administrators, decision-makers, and legislators that if we are to learn anything from experience it is necessary to hold the design of the program (that is, the designed project differentials in treatment variables) constant for a long enough period of time to allow for the "settling down" of the program and the collection and analysis of the data. *A commitment to hold to design for a long enough period so that we could learn from experience is a central element in the experimental approach to social action.*

The idea that social action programs should be experimental is simple, but we cannot be sanguine about the speed with which the full implications of this simple idea will be accepted by decision-makers and the public as a whole. The view that programs can be large scale *action* programs and still be designed as intentional experiments has not been easy to get across, even to those trained in experimental methods in the social sciences, with its tradition of small scale research.

The emphasis on ex post evaluation is evidence of the fact that at some level legislators understand that social action programs are "testing" concepts. But it will require more explicit acceptance of the idea that some aspects of programs "tested" in action will fail before the full advantages of the intentionally experimental approach can be realized. It takes restraint to mount a program with a built-in experimental design and wait for it to mature before deciding on a single program concept, but we emphasize that restraint does not mean small scale or limited action.

It is not unfair, we think, to characterize the approach to social action programs that has been taken in the past as one of serial experimentation through program failure. A program is built around a single concept, eventually it is realized that it does not work, so the program is scrapped (or allowed to fade away) and a new program and concept is tried. Certainly serial experimentation through failure is the hard way to learn. An intentionally experimental approach would allow us to learn faster by trying alternative concepts *simultaneously* and would make it more likely that we could determine not only *that* a particular concept failed, but also *why* it failed.

The Acceptability of Evaluation Results

It does little violence to the facts to state that few decisions about social action programs have been made on the basis of the types of evaluations we have been discussing thus far in this paper. A major reason for this, we feel, is an inadequate taste for rigor (or an overweening penchant for visceral judgments) by administrators and legislators and excessive taste for the purely scientific standards by academics. It often seems that the scholars conspire with the legislators to beat down any attempt to bring to bear more orderly evidence about the effectiveness of alternative programs; it is not at all difficult to find experts who will testify that virtually any evaluation study is not adequately "scientific" to provide a sound basis for making program decisions. There is a reasonable

and appropriate fear on the part of academics that sophisticated techniques of analysis will be used as deceptive wrapping around an essentially political kernel to mislead administrators or the public. This fear, however, often leads to the setting of standards of "proof" which cannot, at present, given the state of the art of social sciences, or perhaps never, given the inherent nature of social action programs, be satisfied. The result generally is that the evaluation is discredited, the information it provides ignored, and the decision-maker and legislator can resume the exercise of their visceral talents.

A first step toward creating a more favorable atmosphere for evaluation studies is to recognize that they will not be final arbiters of the worth of a program. A positive but more modest role for evaluation research was recently stated by Kenneth Arrow in a discussion of the relative virtues of the tradition processes of public decision-making (characterized as an adversary process) and the recently developed procedure of the Programming, Planning, Budgeting System (characterized as a rationalistic or "synoptic process".[16] Arrow advocated an approach in between forensics and synoptics.[17] He illustrated his argument by making an analogy with the court system, suggesting that what was happening through the introduction of the more rationalistic processes was the creation of a body of "rules of evidence." The use of systematic evaluation (along with the other elements of the PPBS) represents an attempt to raise the standards of what is admissible as evidence in a decision process that is inherently likely to remain adversary in nature. Higher standards of evaluation will lessen the role of "hearsay" testimony in the decision process, but they are not meant to provide a hard and fast decision rule in and of themselves. The public decision-making process is still a long way from the point at which the evidence from a hard evaluation is the primary or even the significant factor in the totality of factors which determine major decisions about programs. Therefore, the fear of many academics that poorly understood evaluations will ex-

ercise an inordinate influence on public decisions is, to say the least, extremely premature. But if standards for the acceptance of evaluation results are viewed in terms of the "rules of evidence" analogy, we can begin to move toward the judicious mix of rigor and pragmatism that is so badly needed in evaluation analysis.

The predominant view of the role of "serious," independent evaluations [18] (particularly in the eyes of harried administrators), seems to be that of a trial (to continue the analogy) aimed at finding a program guilty of failure. There is a sense in which this paranoid view of evaluation is correct. The statistical procedures used usually start with a null hypothesis of "no effect," and the burden of the analysis is to provide evidence that is sufficiently strong to overturn the null hypothesis. As we have pointed out, however, problems of data, organization, and methods conspire to make clear-cut positive findings in evaluations difficult to demonstrate.

The atmosphere for evaluations would be much healthier if the underlying stance were shifted from this old world juridicial rule. Let the program be assumed innocent of failure until proven guilty through clear-cut negative findings. In more precise terms, we should try to avoid committing what are called in statistical theory Type II errors. Thus, an evaluation which does not permit rejecting the null hypothesis (of a zero effect of the program) at customary levels of statistical significance, may be consistent with a finding that a very large positive effect may be just as likely as a zero or negative effect.[19] "Rules of evidence" which emphasize the avoidance of Type II errors are equivalent to an attitude which we have characterized as "innocent until proven guilty." (We must frankly admit that, like court rules of evidence, this basic stance may provide incentives to the program administrators to provide data which are sufficient only for arriving at a "no conclusion" evaluative outcome.)

As a final conciliatory comment; when we talk about evaluation studies leading to verdicts of "success" or "failure," it should be recognized that we are greatly simplifying and abbreviating the

typical results. Most social action programs are so complex in the variety of inputs and the multiplicity of objectives, that simple over-all judgments are not likely to lead to quick decisions to dump programs. In combination with more detailed studies, the purpose of the evidence provided by the analysts will instead usually be to suggest modifications in the program—to shift the composition of inputs, perhaps to re-emphasize some objectives and de-emphasize others—and to suggest marginal additions or subtractions in the total scale of the program. It is worth emphasizing these modest objectives because the trust and cooperation of program administrators are indespensable to an evaluation of the program.

Footnotes

[1] As examples of the benefit-cost literature, see Robert Dorfman, ed., *Measuring Benefits of Government Investments* (Brookings Institution, Washington, D.C., 1965), and A. R. Prest and R. Turvey, "Cost-Benefit Analysis: A Survey," *Economic Journal*, December, 1965, v. 75, pp. 683–735. As examples of the evaluation research literature, see Edward A. Suchman, *Evaluation Research* (Russell Sage Foundation, New York, 1967), Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago, Rand-McNally, 1966), G. H. Orcutt and A. G. Orcutt, "Incentive and Distincentive Experimentation for Income Maintenance Policy Purposes," *American Economic Review*, September, 1968, v. 58, pp. 754–72, and Harold Watts, "Graduated Work Incentives: Progress toward an Experiment in Negative Taxation," Discussion Papers Series, Institute for Research on Poverty, University of Wisconsin, 1968. For examples of the point of view of officials of governmental agencies, see William Gorham, "Notes of a Practitioner," and Elizabeth Drew, "HEW Grapples with PPBS," in *The Public Interest*, Summer, 1967, No. 8.

[2] There does seem to be a developing literature in which the a priori benefit-cost estimates are compared with the ex post results for water projects. See Maynard Hufschmidt, " 'Systematic Errors' in Cost Estimation in Public Investment," to appear in the Universities-National Bureau of Economic Research Conference volume, *The Economics of Public Output*. It may be that similar follow-up studies are being undertaken for defense projects—one can at least say that Congressional committees are determined to carry out their own follow-up evaluations on projects such as the TFX.

[2a] In a more extended version of this paper prepared for the Institute for Research on Poverty, University of Wisconsin (Discussion Paper 42–69), we discuss several problems associated with the specification of objectives of programs and how these affect evaluation designs.

[3] We are indebted to Thomas K. Glennen, RAND Corporation, for his ideas on this point.

[4] Briefly, the OIC concept combines elements of training, job development (often aided by pressure tactics against employers), and a psychological up-lifting of the participants which is conducted with an ideology of militancy and participatory democracy.

[5] The Work Experience program consisted of public employment of welfare recipients and other adult poor under Title V of the Economic Opportunity Act. Only minimal training was offered, but it was hoped that work-for-pay would, by itself, provide a springboard to self-sustaining employment in the private market.

[6] U. S. Congress, House Committee on Ways and Means, *Community Work and Training Program.* 90th Congress, 1st Sess., House Document No. 96 (Washington, D.C.: U. S. Government Printing Office, 1967).

[7] Worth Bateman, "Assessing Program Effectiveness," *Welfare in Review,* Vol. 6, No. 1, January-February 1968.

[7a] See the version of this paper in the Discussion Series of the Institute for Research on Poverty.

[8] An important point to be remembered is that, for any given amount of resources available for an evaluation study, there is a trade-off between an allocation of these resources for increased sample size and allocation for improved quality of measurement, which might take the form of an expanded set of variables, improved measures of variables, or reduced attrition from the sample. Too often we have witnessed a single-minded attachment to larger sample sizes, probably stemming from the analyst's fear that he will end up with "too few observations in the cells" of some only vaguely imagined cross-tabulation. This fear should be balanced by an awareness both of the rapidity with which marginal gains in precision of estimates decline with increases in "medium size" samples and of the extent to which a theoretically justified multiple regression model can overcome some of the limitations which cross-tabulation analysis impose on a given-sized sample.

[9] See the vigorous defense of an experimental method in social action programs in: Guy H. Orcutt and Alice G. Orcutt, *op. cit.* ....

[10] This assumption will strike some readers as too positivistic, too restrictive to "things measurable," and too oblivious to the unmeasurable and subjective variables. Let us say in defense of this assumption only that it is a "working assumption" that permits us to discuss an important region of evaluation which covers the measurable portion, that it is desirable to expand this region and, therefore, to narrow the area left for subjective judgments, and that, in any case, the objective portion is necessary to an improved over-all judgment that spans both measurable and unmeasurable inputs and outputs of a program.

[10a] We bypass here the important question of the choice of a discount rate. Some discussion of this issue is provided in the Institute for Research on Poverty version of this paper.

[11] When the program produces an increase in consumption of goods and services, the treatment of these transfer payments can become more complicated if we do not assume that the goods and service have a value to the recipients equal to their cost. See A. A. Alchian and W. R. Allen, *University Economics* (Wadsworth: Belmont, California, 1967, Second Edition) pp. 135–140 for an extended discussion.

[12] For just one of many examples of this type of treatment of transfer payments see, "The Feasibility of Benefit-Cost Analysis in the War on Poverty: A Test Application to Manpower Programs," prepared for the General Accounting Office, Resource Management Corporation, UR-054, December 13, 1968.

[13] For a notable exception to the absence of attempted measurement of the type of third-party discussed above, see Thomas I. Ribich, *Education and Poverty* (Washington, D.C.: The Brookings Institution, 1968). Ribich's study also gives us some evidence of the likelihood of relatively small quantitative magnitudes of these effects. A rather free wheeling listing of third-party effects runs the risk of double counting benefits. For example, although other family members benefit from the better education or earnings of the head of the household, we should not forget that had the investment expenditure been elsewhere, even if in the form of an across-the-board tax cut, *other* family heads would have had larger incomes, at least, with resulting benefits to *their* families. In his examination of cost-benefit analysis of water resources developments, Roland N. McKean gives an extended discussion of the pitfalls of double counting. See his *Efficiency in Government Through Systems Analysis* (New York: John Wiley and Sons, Inc., 1958), especially Chapter 9.

[14] An exceptionally good discussion of negative external effects, including disruption

of the community structure, is contained in Anthony Downs, "Uncompensated Non-Construction Costs Which Urban Highways and Urban Renewal Impose on Residential Households" which will appear in a Universities-National Bureau of Economic Research Conference volume entitled, *Economics of Public Output*. The literature on urban renewal and public housing is extensive and too well known to require listing here.

[15] For an excellent discussion of many of these issues see Joel F. Handler, "Controlling Official Behavior in Welfare Administration," *The Law of the Poor*, ed., J. tenBroek (Chandler Publishing Co., 1966). (Also published in *The California Law Review*, Vol. 54, 1966, p. 479.)

[16] For a more complete discussion of this terminology, see Henry Rowen, "Recent Developments in the Measurement of Public Outputs," to be published in a Universities-National Bureau of Economic Research Conference volume, *The Economics of Public Output*.

[17] Remarks by Kenneth Arrow during the NBER conference cited in the previous footnote.

[18] We mean here to exclude the quick and casual sort of evaluations, mainly "in-house" evaluations, that more often than not are meant to provide a gloss of technical justification for a program.

[19] Harold Watts has stressed this point in conversations with the authors. See Glen G. Cain and Harold W. Watts, "The Controversy about the Coleman Report: Comment," *Journal of Human Resources*, Vol. III, No. 3, Summer, 1968, pp. 389–92, also, Harold W. Watts and David L. Horner, "The Educational Benefits of Head Start: A Quantitative Analysis," Discussion Paper Series, The Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin.

PUBLICATIONS OF THE INDUSTRIAL RELATIONS RESEARCH INSTITUTE AND CENTER FOR STUDIES IN VOCATIONAL AND TECHNICAL EDUCATION

State Labor and Social Legislation. A Symposium in Honor of Elizabeth Brandeis Raushenbush, May 20, 1966. $4.50 cloth, $2.50 paper.

†Archie Kleingartner. *Professionalism and Salaried Worker Organization.* 1967.

†Kenneth McLennan. *Managerial Skill and Knowledge.* 1967.

The Labor Movement: A Re-examination. A Conference in Honor of David J. Saposs, January 14 and 15, 1966. $2.50.

Bhal J. Bhatt. *Labor Market Behavior of Factory Workers in Bombay.* 1969. $4.50 cloth, $3.00 paper.

CONFERENCE PROCEEDINGS OF THE CENTER

†*Occupational Data Requirements for Education Planning,* June 15 and 16, 1965.

Research in Vocational and Technical Education, June 10 and 11, 1966. $3.00.

Research in Apprenticeship Training, September 8 and 9, 1966. $3.00.

†*Research Approaches to the Initiation of New Vocational-Technical Programs,* October 3-7, 1966.

The Encouragement of Research in Predominantly Negro Universities, February 23, 1967.

The Education and Training of Racial Minorities, May 11 and 12, 1967. $3.00.

Education and Training in Correctional Institutions, 1968. $3.00.

PERIODICALS

The Journal of Human Resources, published quarterly. $8.00 per year for individuals, $12.00 per year for institutions (add 50 ¢ a year for foreign mailing). Subscriptions available from Journals Department, The University of Wisconsin Press, P. O. Box 1379, Madison, Wisconsin 53701.

IRRI Report, published semiannually.

CENTER REPRINTS*

†W. Lee Hansen. "Human Capital Requirements for Educational Expansion: Teacher Shortages and Teacher Supply," *Education and Economic Development* (1965).

†William J. Swift and Burton A. Weisbrod, "On the Monetary Value of Education's Intergeneration Effects," *Journal of Political Economy* (December 1965).

†W. Lee Hansen. "Labor Force and Occupational Projections," *Proceedings of the 18th Annual Meeting* of the Industrial Relations Research Association, 1965.

†Burton A. Weisbrod. "Investing in Human Capital," *Journal of Human Resources* (Summer 1966).

†Gerald G. Somers. "Government-Subsidized On-The-Job Training: Surveys of Employer's Attitudes," *Hearings before the Subcommittee on Employment and Manpower,* 89th Congress (September 1965; February 1966).

†Burton A. Weisbrod. "Conceptual Issues in Evaluating Training Programs," *Monthly Labor Review* (October 1966).

†Gerald G. Somers. "Retraining the Unemployed Older Worker," *Technology, Manpower and Retirement Policy.* (1966).

†Glen Cain, W. Lee Hansen, and Burton A. Weisbrod. "Occupational Classification: An Economic Approach," *Monthly Labor Review* (February 1967).

Norman F. Dufty, "Apprenticeship—A Theoretical Model," *British Journal of Industrial Relations,* (March 1967).

†J. K. Little. "The Occupations of Non-College Youth," *American Educational Research Journal,* (March 1967).

†Glen Cain and Gerald Somers. "Retraining the Disadvantaged Worker," *Research in Vocational and Technical Education,* Proceedings of a Conference, 1966.

Gerald G. Somers. "Our Experience With Retraining and Relocation," *Toward a Manpower Policy* (1967).

†Jack Barbash. "Union Interests in Apprenticeship and Other Training Forms," *Journal of Human Resources* (Winter 1968).

Gerald G. Somers. "The Response of Vocational Education to Labor Market Changes," *Journal of Human Resources,* Supplement (1968).

†Walter Fogel. "Labor Market Obstacles to Minority Job Gains," *Proceedings of the 20th Annual Winter Meeting of the Industrial Relations Research Association,* 1967.

Gerald G. Somers and Graeme H. McKechnie. "Vocational Retraining Programs for the Unemployed," *Proceedings of the 20th Annual Winter Meeting of the Industrial Relations Research Association,* 1967.

Burton A. Weisbrod and Peter Karpoff. "Monetary Returns to College Education, Student Ability, and College Quality," *Review of Economics and Statistics* (November 1968).

David B. Johnson and James L. Stern. "Why and How Workers Shift From Blue-Collar to White-Collar Jobs," *Monthly Labor Review,* (October 1969).

Philip A. Perrone and Donald H. Johnson. "The Marginal Worker: Projections of High School Vocational Teachers," *Journal of Human Resources* (Fall 1968).

Glen G. Cain and Robinson G. Hollister. "The Methodology of Evaluating Social Action Programs," *Public-Private Manpower Policies,* Industrial Relations Research Association (November 1969).

†Out of Print
*Single copies available without charge

IRRI REPRINTS 1967 TO THE PRESENT

(Reprints prior to 1967 are not listed. A complete listing of IRRI reprints may be obtained by writing to the IRRI, 4315 Social Science Building, 1180 Observatory Drive, Madison, Wisconsin 53706.)

80 "Current Research Activities of the Industrial Relations Research Institute, 1966," *Industrial and Labor Relations Review* (January 1967).
† 81 James Stern. "Adjustment to Plant Closure," *Monthly Labor Review* (January 1967).
† 82 Richard U. Miller. "Arbitration of New Contract Wage Disputes: Some Recent Trends," *Industrial and Labor Relations Review* (January 1967).
† 83 Glen Cain. "Unemployment and the Labor-Force Participation of Secondary Workers," *Industrial and Labor Relations Review* (January 1967).
† 84 Alan C. Filley. "Behavioral Science Applications to Management: An Appraisal," *Business Perspectives* (Winter 1966).
85 Everett M. Kassalow. "National Wage Policies: Lessons to Date, Europe and the U.S.A.," *Proceedings to the 19th Annual Winter Meeting of the Industrial Relations Research Association, 1966.*
† 86 Gerald G. Somers. "The Rich, The Poor, and The Others," *The Changing American Economy* (1967).
† 87 W. Lee Hansen. "The Economics of Scientific and Engineering Manpower," *Journal of Human Resources* (Spring 1967).
† 88 Jack Barbash. "Technology and Labor in the Twentieth Century," *Technology in Western Civilization* (1967).
† 89 W. Lee Hansen and Burton A. Weisbrod. "Economics of the Military Draft," *Quarterly Journal of Economics* (August 1967).
† 90 Jack Barbash. "The Structure and Evolution of Urban Interests in Pensions," Joint Economic Committee, Subcommittee on Fiscal Policy, *Old Age Income Assurance Part IV* (December 1967).
† 91 Everett M. Kassalow. "Trade Unionism and the Development Process in the New Nations: A Comparative View," *International Labor* (1967).
† 92 Jack Barbash. "John R. Commons and the Americanization of the Labor Problem," *Journal of Economic Issues* (September 1967).
† 93 Norman F. Dufty. "Blue Collar Contract," *International Journal of Comparative Sociology.*
94 Richard U. Miller. "Labor Legislation and Mexican Industrial Relations," *Industrial Relations* (February 1968).
† 95 David Chaplin. "Industrial Labor Recruitment in Peru," *America Latina* (October –December 1968).
† 96 George Huber, "The Application of Behavioral Science Theory to Professional Development," *Academy of Management* (September 1967).
† 97 Gerald Somers and Burton Weisbrod. "The Development of Human Resources," *Federal Programs for the Development of Human Resources,* Joint Economic Committee, U.S. Congress (1968).
† 98 Raymond Munts and Mary Louise Munts. "Welfare History of the the I.L.G.W.U.," *Labor History,* Supplement (Spring 1968).
† 99 Jack Barbash. "The I.L.G.W.U. as an Organization in the Age of Dubinsky," *Labor History,* Supplement (Spring 1968).
†100 Emery F. Via. "Discrimination, Integration, and Job Equality," *Monthly Labor Review* (March 1968).
†101 David Chaplin. "Labour Turnover in the Peruvian Textile Industry," *British Journal of Industrial Relations* (March 1968).
†102 A. J. Grimes. "Personnel Management in the Building Trades," *Personnel Journal* (January 1968).
103 "Current Research Activities of the Industrial Relations Research Institute, 1968," *Industrial and Labor Relations Review* (July 1968).
†104 A. C. Filley and A. J. Grimes. "The Bases of Power in Decision Processes," *Academy of Management Proceedings* (December 1967).
†105 Jack Barbash. "American Unionism: From Protest to Going Concern," *Journal of Economic Issues* (March 1968).
†106 Everett M. Kassalow. "The Comparative Labour Field," Bulletin No. 5, International Institute for Labour Studies (November 1968).
†107 James L. Stern. "Alternative Dispute Settlement Procedures," *Wisconsin Law Review.*
108 Everett M. Kassalow. "Professional Unionism in Sweden," *Industrial Relations* (February 1969).
109 Raymond Munts. "Collective Bargaining for Health as a Form of Health Planning," Public Health Service–Labor Seminar on *Consumer Health Services,* 1968.
†110 Everett M. Kassalow. "Trade Unionism Goes Public," *The Public Interest* (Winter 1969).
111 Harvey A. Juris. "Union Crisis Wage Decisions," *Industrial Relations* (May 1969).
112 George Hagglund. "Why Industrial Engineers Are Needed in Agriculture," *Industrial Engineering* (April 1969).
113 Robert J. Lampman. "Some Interactions Between Economic Growth and Population Change in the Philippines" and "The Sources of Post-War Economic Growth in the Phillippines," *Phillippine Economic Journal.*
114 Robert J. Lampman. "The Investment of Social Security Reserves and Development Problems: The Philippines as a Case History," *The Role of Social Security in Economic Development,* Research Report No. 27.
115 Gerald G. Somers. "Data Needs for Monitoring and Evaluating Manpower Programs," *Proceedings of the 21st Annual Winter Meeting of the Industrial Relations Rearch Association,* 1968.
116 Everett M. Kassalow. "Public Employee Bargaining in Europe: What Lessons for the United States?" *Proceedings of the 21st Annual Winter Meeting of the Industrial Relations Research Association,* 1968.
117 W. Lee Hansen and Burton A. Weisbrod. "The Distribution of Costs and Direct Benefits of Public Higher Education: The Case of California," *Journal of Human Resources* (Spring 1969).
118 Gerald G. Somers. "Bargaining Power and Industrial Relations Theory," *Essays in Industrial Relations,* Ch. 3.
119 Jack Barbash. "Rationalization in the American Union," *Essays in Industrial Relations,* Ch. 10.
120 Murray Edelman. "The Conservative Political Consequences of Labor Conflicts," *Essays in Industrial Relations,* Ch. 11.
121 Stanley Sloan and Alton C. Johnson. "Performance Appraisal: Where Are We Headed?" *The Personnel Administrator* (September–October 1969).
122 Raymond Munts. "The New Realities of Unemployment Insurance: Objectives for Today's Economy," *Manpower* (August 1969).
123 Stanley Sloan. "Democracy in a Public Employee Union," *Public Personnel Review* (October 1969).
124 David Chaplin. "Blue-Collar Workers in Peru," *International Journal of Comparative Sociology* (March and June 1969).

ERIC Clearinghouse
DEC 15 1970
on Adult Education